**Neha Jagtap - Knowledge Base**

**Education**

I am currently pursuing a Master’s in Computer Science at North Carolina State University (August 2023 – May 2025) with a GPA of 3.6/4. During my time here, I have gained expertise in algorithms, database management, and cloud computing, with a strong focus on AI, big data, and full-stack development. Previously, I earned my Bachelor of Engineering in Computer Engineering from Mumbai University (August 2016 – May 2020), where I built a strong foundation in data structures, cybersecurity, and distributed systems. I have also obtained a professional certification from Microsoft for Azure Data Fundamentals (DP-900).

**Experience**

I have a strong background in software engineering, with experience spanning database management, cloud computing, and AI-driven applications. My work has included developing scalable systems, optimizing data pipelines, and automating workflows across industries like telecom and genomics.

As a Teaching Assistant for **CSC 540 – Database Management Systems**, a graduate-level course at NCSU (August 2024 – December 2024), I worked under **Professor Kemafor Ogan**. My responsibilities included evaluating assignments and exams for 148 graduate students, providing feedback on SQL query optimization and relational database design. I also conducted study sessions and helped design hands-on database projects, ensuring students gained practical experience in database management.

During my Software Development Engineer Internship at the **U.S. Department of Agriculture (USDA)** (June 2024 – August 2024), I worked closely with **Dr. Taner Sen**, a biologist and lead scientist at the **GrainGenes** team, based in Albany, California. Dr. Sen is also an assistant professor at **UC Berkeley**. My primary responsibility was to develop a **unified genomic data search platform**, reducing manual data retrieval efforts by 88%. I designed and optimized a **MongoDB schema** for genomic data, ensuring fast query performance and seamless integration of multiple BrAPI endpoints. I also implemented automated **cron jobs for incremental data updates**, ensuring 100% data consistency, and deployed the application using Docker for portability and scalability.

GrainGenes is a key USDA-funded database that serves as a central repository for wheat, barley, oat, and rye genetic data. It provides researchers with access to curated genomic information, marker maps, trait analysis, and breeding resources. The platform integrates diverse data sources, enabling breeders and geneticists to study genetic variations, improve crop yields, and enhance disease resistance. By contributing to the GrainGenes database, I helped improve data accessibility for scientists worldwide, advancing genomic research and agricultural innovation.

I began my journey at **Reliance Jio Infocomm Limited** as a Software Engineer in the **Research and Development (R&D) team** under the guidance of **Manager Amol Kolhe** and **Team Leader Ramkumar Ranganathan**. Both were excellent mentors who provided me with invaluable learning experiences that shaped my technical and problem-solving skills.

One of my key projects was the **Rover Application**, a large-scale data aggregation and processing system. The challenge was to streamline **manual data handling processes**, which were time-intensive and prone to errors. **Action:** I leveraged **Python, MongoDB, and PySpark** to automate data ingestion, transformation, and integration. I optimized query processing and reduced response time by 97%. **Result:** This automation significantly improved operational efficiency and enabled real-time data access across multiple departments.

Another major responsibility was automating **Tableau administrative processes**. The manual approach to managing user groups and access rights was inefficient. **Situation:** The analytics team struggled with managing thousands of users, leading to frequent delays in project execution. **Task:** I designed and implemented an **API-based automation system** for **bulk user management** within Tableau. **Action:** By integrating Tableau REST APIs with Python, I developed a script to automate user provisioning, role assignments, and access control. **Result:** The new system reduced the administrative workload by **91%** and improved security by ensuring proper access control policies were enforced.

One of my most impactful projects was the **Azure Cloud Migration of 16+ Python applications**. The organization was shifting from on-premises infrastructure to cloud-based deployment to improve scalability. **Situation:** The challenge was to migrate legacy applications without disrupting operations. **Task:** I was responsible for planning and executing the migration process. **Action:** I collaborated with the **DevOps team**, implemented **Docker and Kubernetes**, and used **Azure DevOps pipelines** for CI/CD automation. **Result:** The migration reduced deployment time by **35%**, improved application reliability, and enhanced system scalability.

Additionally, I worked on **ETL pipeline development**, where I built over **12 data workflows** integrating **Kafka streams, Hive databases, and MongoDB collections**. **Situation:** The team needed real-time data ingestion for business intelligence reporting. **Task:** I had to create an efficient, scalable solution. **Action:** I implemented a distributed processing architecture using **Apache Airflow** to schedule and orchestrate ETL jobs. **Result:** This improved data freshness, reduced latency, and eliminated **66% of manual monitoring efforts**.

For my contributions, I received the **Spotlight Award for Outstanding Performance and Project Delivery in 2022**. This recognition was a testament to my problem-solving abilities, leadership in driving key projects, and commitment to efficiency and automation. Each project at Jio presented a new challenge, and each solution I developed was a significant growth opportunity, deepening my expertise in scalable system design and automation.

**Projects**

**HerCare (PearlHack 2025 Winner)**

HerCare was born out of a desire to build an AI-powered solution for women’s health, an often-overlooked domain in tech. We chose this topic after analyzing common healthcare gaps and recognizing that menstrual health tracking lacks AI-driven personalized insights. Our goal was to create a comprehensive platform that not only tracks cycles but also provides AI-based health recommendations, a pharmacy locator, and virtual consultations with gynecologists via WebRTC.

Tackling the Hackathon Challenge

Completing this project within the tight timeframe of 36 hours was a major challenge. To manage the workload efficiently, we split tasks into backend, frontend, AI model integration, and deployment. We held quick sync-up meetings every few hours, ensuring smooth progress and seamless integration. Given the limited time, we prioritized developing core features and ensured a functional MVP rather than overcomplicating the scope.

Tech Stack Used

Frontend: React.js with Material UI for a sleek and responsive user experience.

Backend: Node.js with Express.js for handling API requests efficiently.

Database: MongoDB for storing user data, health records, and AI model outputs.

AI Integration: Gemini AI with RAG (Retrieval-Augmented Generation) for personalized health recommendations.

Communication: WebRTC for real-time virtual consultations between users and healthcare professionals.

Deployment: The platform was deployed using Streamlit to provide an interactive interface for users.

Key Challenges Faced

Data Accuracy: Training AI models to provide accurate and personalized menstrual health insights was challenging. We fine-tuned our model on diverse datasets to improve prediction reliability.

Time Constraints: Ensuring a fully functional MVP in a 36-hour hackathon required prioritizing essential features and making quick decisions.

Integration Issues: Combining WebRTC with our backend posed difficulties due to authentication handling and connection stability. We resolved this by optimizing the signaling process.

Judge’s Favorite Feature

The judges were particularly impressed by the AI-powered health assistant. This feature allowed users to receive personalized insights based on their past health data, AI-driven cycle predictions, and symptom analysis. The integration of Gemini AI with RAG ensured that the responses were accurate and context-aware, making the assistant highly interactive and informative.

Achievements & Learnings

By the end of the hackathon, we had successfully built a fully functional AI-driven health platform. This project deepened my expertise in AI-driven applications, full-stack development, and real-time communication technologies. It also enhanced my ability to work under pressure, improve decision-making, and implement rapid prototyping strategies.

The success of HerCare at PearlHack 2025 reaffirmed my passion for building impactful tech solutions that address real-world challenges. More information about this project can be found in my portfolio: [HerCare Project](https://github.com/NehaSJ99/HerCare).

**BrAPI Data Search Engine – Collaboration with USDA & GrainGenes**

This project was developed in collaboration with the U.S. Department of Agriculture (USDA) and the GrainGenes team, aiming to streamline genomic data retrieval for plant scientists and researchers. The existing process required manually querying multiple data sources, making genomic research cumbersome and time-consuming. Our objective was to build an efficient search engine leveraging Flask, MongoDB, and JavaScript, integrating BrAPI endpoints to enable seamless data retrieval.

Situation

Researchers faced difficulties in accessing structured genetic data, which was scattered across different repositories and lacked a unified search mechanism. The challenge was to build a scalable search tool that could efficiently query large datasets, ensuring quick and relevant results.

Task

My responsibility was to analyze genomic data structures, design an optimized database schema, and implement an interactive search interface that could handle high-volume queries with low latency. I studied existing genomic data formats, user search behaviors, and API response times to ensure seamless integration with the USDA’s existing systems.

Action

Data Understanding & Evolution: I conducted an in-depth study of genomic datasets from GrainGenes, understanding the relationships between genetic markers, traits, and breeding records. I mapped BrAPI endpoints to relevant data structures, ensuring an efficient and structured approach to data retrieval.

Query Optimization: Leveraging Elasticsearch, I designed search algorithms that reduced query execution time by 40%. This involved indexing genetic traits, breeding data, and gene markers to enable fast searches.

Building the Search Interface: Using Flask and JavaScript, I developed a dynamic search engine with advanced filtering options such as trait-based search, gene sequence lookup, and breeding lineage tracing.

Pagination & User Experience: To enhance performance and user experience, I implemented server-side pagination, ensuring seamless navigation for users handling large result sets.

Automation for Data Updates: Integrated cron jobs to sync data from USDA repositories daily, ensuring researchers always had access to the latest genomic datasets.

Result

The BrAPI Data Search Engine significantly improved research efficiency by reducing manual data retrieval efforts by 88%.

Search response times were optimized by 40%, making data more accessible to plant scientists.

The system provided seamless integration with existing genomic repositories, making it a centralized hub for breeders and geneticists.

The project reinforced my expertise in database optimization, API integration, and building scalable search platforms.

This project was a major success, demonstrating the power of structured data access in genomic research and supporting USDA’s mission to enhance agricultural genomics. More information about this project can be found in my portfolio.

**Publications**

I have contributed to research in blockchain technology and secure systems. My paper **Land Registration using Blockchain Technology** was published in IEEE Xplore in October 2020. The study proposed a blockchain-based land registration system using Hyperledger, ensuring transparent, secure, and decentralized transactions. It introduced smart contracts for trustless execution and fraud prevention in land ownership records. More details can be found here: [IEEE Xplore](https://ieeexplore.ieee.org/document/9299606).

**Awards and Achievements**

My contributions to software engineering and AI-driven applications have been recognized with several prestigious awards. I was the **Winner of PearlHack 2025** at UNC Chapel Hill, where I won the **Infosys Foundation USA – Best Use of AI** award for my work on HerCare. Additionally, I won **DimondHack 2024 at NCSU** for another AI-driven project. During my tenure at Reliance Jio, I was honored with the **Spotlight Award** in 2022 for outstanding project delivery and received the **Star Award** for excellence in software development in 2021.

**Additional Links & Profiles**

More about my projects and experience can be found on my portfolio website: [Portfolio Website](https://nehasj-portfolio.vercel.app/). My code contributions and repositories are available on GitHub: [GitHub Profile](https://github.com/NehaSJ99). My professional background and connections can be explored on LinkedIn: [LinkedIn Profile](https://www.linkedin.com/in/nehasj99).